DOCUMENT RESUME

ED 442 210 EC 307 845

AUTHOR Huesman, Ronald L., Jr.; Frisbie, David A.

TITLE The Validity of ITBS Reading Comprehension Test Scores for Learning Disabled and Non Learning Disabled Students under Extended-Time Conditions.

PUB DATE 2000-04-00

NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Academic Achievement; Grade 6; Intermediate Grades; *Learning Disabilities; *Reading Comprehension; *Test Validity; Testing; Time Factors (Learning); *Timed Tests

IDENTIFIERS *Iowa Tests of Basic Skills; *Testing Accommodations (Disabilities)

ABSTRACT

This study investigated the effect of extended-time limits in terms of performance levels and score comparability for reading comprehension scores on the Iowa Tests of Basic Skills (ITBS). The first part of the study compared the average reading comprehension scores on the ITBS of 61 sixth-graders with learning disabilities and 397 non learning disabled sixth-graders under two timing conditions (extended-time vs. standard-time). The second part of the study examined whether the scores arising from the two timing conditions measured the same construct (i.e., reading comprehension) for both groups of students. Results indicate that students with learning disabilities made significantly larger gains on the ITBS Reading Comprehension Test under extended-time conditions than students without learning disabilities who received appropriate timing instructions. Results also found testing directions had an effect on student performance. Non learning disabled students given instructions to take their time did not perform any differently than students with learning disabilities under extended-time conditions. When students without disabilities were told to work at a slow and careful pace, where time was not a factor, they made significant gains under extended-time conditions; however, when they were told to work at a normal rate, they did not make significant gains. (Contains 33 references.) (CR)

Reproductions supplied by EDRS are the best that can be made
from the original document.

THE VALIDITY OF ITBS READING COMPREHENSION TEST SCORES FOR LEARNING DISABLED AND NON LEARNING DISABLED STUDENTS UNDER EXTENDED-TIME CONDITIONS

ED 442 210

Ronald L. Huesman, Jr.
David A. Frisbie
The University of Iowa

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Huesman, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper presented at the Annual Meeting
of the National Council on Measurement in Education
New Orleans, LA
April, 2000

EC 307845

BEST COPY AVAILABLE

Abstract

The purpose of this study was to determine the effect of extended-time limits on test performance and score comparability for the ITBS Reading Comprehension scores of learning disabled (LD) and non learning disabled (NLD) students. The extension of testing time is expected to alleviate an irrelevant source of difficulty for LD students (i.e., slower rate of information processing) and allow them enough time to demonstrate their achievement.

Students identified by their school as LD (n=129) and two groups of NLD students (n=235 and n=162) participated. The two NLD groups were not combined due to unplanned differences in test administration conditions. Testing occurred as part of each school's annual testing program and was to resemble a standard administration, except for the extension of time limits. At the end of the standard-time limit, students marked on their answer folder the last item answered. If any student had not completed the test by the end of the standard time limit, additional 20-minute blocks of time were given.

For NLD students given directions to work at a normal rate, there was little difference in test performance between timing conditions, and a factor analysis using passage-based testlet correlations found a one-factor model fit the data from both timing conditions. For NLD students who were told to take their time and work carefully, test performance increased with added time, and the factor structure differed between timing conditions. A two-factor model fit better under standard-timing conditions, and a one-factor model fit better under extended-time conditions. A similar result occurred for LD students: test performance significantly increased with more time and evidence suggested the factor structures differed, though not to the same degree as for the latter NLD group.

The amount of extra time needed by LD students varied greatly among them, and many finished within the standard-time limit. In addition, the different pacing in the directions given the two groups of NLD students made a difference in their work rate, test performance, and score meaning. Implications for determining testing accommodations in students' IEPs are discussed.

Introduction

The Individuals with Disabilities Education Act of 1991 (IDEA, 1991) was intended to provide educational services for individuals with disabilities and guarantee a free public education for these students (Phillips, 1994). The IDEA Amendments of 1997 (P.L. 105-17) specifically require that students with disabilities be included in state and district-wide assessments and be given appropriate accommodations when necessary. The amendments also require that alternate assessments be provided for students whose IEPs specify that they should be excluded from regular assessments. In addition, states and local districts are required to report on the participation and performance of students with disabilities in the same detail and frequency as students without disabilities. The inclusion of disabled students in the large-scale measurement of achievement should result in more representative state-wide or district-wide samples and, therefore, more valid indicators of all students' achievement. The inclusion of students with disabilities who require accommodations along with non-disabled students in the assessment process may be more fair and desirable, but it also introduces significant score interpretation questions that need to be addressed empirically.

The National Center on Educational Outcomes (NCEO) conducted two of the most comprehensive reviews of the literature on testing accommodations for students with disabilities in 1993 and again in 1996. The 1993 NCEO review of the literature found very little empirically based research on the effects of test accommodations, and the more recent review found that the situation had changed little. The limited empirical research that had been conducted up to that point had been done primarily with college admission tests (NCEO, Minnesota Report #9, 1996).

A small number of studies have attempted to systematically examine the effects of extended-time versus standard-time limits on the test performance of LD examinees (Alster, 1997; Halla, 1988; Harker, 1991; Hill, 1984; Munger & Loyd, 1991; Perlman, Borger, Collins, Elenbogen, & Wood, 1996; Runyan, 1991a, 1991b). With the exception of Perlman et al. (1996), these studies have included a comparison group of NLD subjects assessed under both standard-time and extended-time conditions. These studies vary greatly in terms of the type of samples used (elementary through

upper level college students), types of tests used (college admission tests, elementary and secondary achievement tests, reading comprehension and algebra tests), and degrees of technical adequacy (generalizability, sampling issues, etc.) See Huesman (1999) for a more detailed review of the testing accommodation literature.

A potential problem regarding the generalizability of LD research results is related to the gender of the student. This in turn is related to whether the samples utilized were identified in the school system or identified by researchers. In general, males are more likely to be identified by school systems for special education services (McDonnell, McLaughlin & Morison, 1997), but those females who are identified tend to have lower IQ scores, are more severely impaired, and have larger aptitude-achievement discrepancies than their male counterparts (Vogel, 1990). The proportion of male to female children identified with learning disabilities is generally quite high, with ratios of 3:1 predicted for reading disabilities by DeFries (1989), and 4:1 for learning disabilities in general by Nass (1993). According to Shaywitz S., Shaywitz B., Fletcher and Escobar (1990), the actual prevalence of males and females with dyslexia (a severe reading disability) may be closer to 1:1, but behavioral factors may have more weight in referrals for special education (e.g., males may be more likely to have Attention Deficit Hyperactivity Disorder in addition to a learning disorder). In most cases, the teacher makes the referral for special education services (Anderson, 1997). Anderson's (1997) review of special education referrals supports this notion: teacher's referrals are heavily influenced by the student's gender; males tend to display more hyperactive or disruptive behavior in the classroom; criteria for placement are often based on male norms; and there is a strong belief in a sex-based etiology of LD. The few extended-time studies conducted up to this point have not assessed the role of gender on test performance; therefore, it was not clear if test performance would be affected differently.

The notion of processing-speed deficits among LD students is of major importance because it provides the justification/validation for allowing LD students extended-time accommodations for testing. The extra time compensates for the slower speed and removes an irrelevant source of

difficulty from most standardized tests involving reading. The slower processing of information for LD students in general is well documented (Runyan, 1991a; Ackerman, Dykman, & Peters, 1977; Badian, 1996; Hasselbring, Goin, & Bransford, 1988; Hayes, Hynd, & Wisenbaker, 1986; Kulak, 1993; Geary & Brown, 1991). Dodd, Griswold, Smith, and Burd (1985) also found evidence that LD children have more difficulty estimating the time duration of general activities, situations, or experiences than NLD children, a difficulty which may have an impact on test performance under standard time limits. Also of interest is the variation in the amount of time needed by LD students. Runyan (1991b) found that the amount of time needed by the LD sample to complete the Nelson Denny Reading Test (NDRT) ranged from an extra four minutes to an extra 29 minutes.

The increase in the number of LD students and the recent legal mandates to include these students in large-scale assessments has made the need to establish the validity of these measurements a more urgent one. Because modified test forms or testing conditions are often necessary for many of these students, the interpretation of the test scores obtained under these accommodated administrations is often ambiguous. The test scores from nonstandard administrations may not accurately reflect the student's abilities that were intended to be measured by the instrument (i.e., the scores may not be comparable in meaning to scores resulting from standard administrations). For example, if LD students are given extra time to complete a test, questions arise about whether the scores based on such an accommodation have the same meaning as the scores of students tested under standard-time limits. In part, there is a question about whether the extra time gives an advantage to the LD student.

The balance between honoring the rights of the disabled test taker and maintaining the validity of the interpretation of their test scores is the core issue of testing with accommodations (Phillips, 1994, 1996). Does the test administered with accommodations measure the same construct as the non-accommodated version, or in other words, do the scores mean the same thing? This question is particularly difficult to answer when the disability affects cognitive functioning (e.g., learning disabilities) because of the confounding of the disability with the academic skills that are

often being measured (Phillips, 1994). The various definitions of learning disabilities, the heterogeneity of disabilities within the learning disability category (i.e., severity and subtypes), plus the various combinations of modifications in test format and testing conditions, have made empirical studies of the effects of test accommodations very difficult to undertake (Willingham, Ragosta, Bennett, Braun, Rock & Powers, 1988).

Purpose. The objective for any accommodation is to,"...provide a test that eliminates, insofar as possible, sources of difficulty that are irrelevant to the skills and knowledge being measured" (Willingham et al., 1988, p.3). The extension of time limits is believed to alleviate an irrelevant source of difficulty for LD students (i.e., slower than usual processing of information) and allow them enough time to demonstrate their knowledge and skills. The purpose of this study was to provide empirical evidence of the effect of extended-time limits in terms of: (1) performance levels and (2) score comparability for reading comprehension scores on the Iowa Tests of Basic Skills (ITBS, Hoover, Hieronymus, Frisbie, & Dunbar, 1994).

The first part of the study compared the average reading comprehension scores on the ITBS of students under two timing conditions (extended-time vs. standard-time). Subgroup breakdowns based on reading ability and verbal ability, also were examined. The dependent measure for the first part of the study was a difference score, obtained by subtracting a student's score under extended-time conditions from the score obtained under standard-time conditions. These research questions were addressed:

1. Is the average difference score for LD students greater than that of the NLD students?
2. Does the relationship between the difference score and the extended-time score vary by overall reading comprehension ability in the same way for LD and NLD students?
3. Is the relationship between the difference scores and gender different within the LD and NLD groups?
4. Is the relationship between the difference scores and ITBS Vocabulary scores different within the LD and NLD groups?

The second part of the study examined whether the scores arising from the two timing conditions measure the same construct (i.e., reading comprehension) for both LD and NLD students. It was hypothesized that, under standard-time conditions, the factor structure might reflect a difficulty factor (speededness) for the LD group that would not be present under the extended-time condition. Under extended-time conditions, if the factor structure is relatively the same for various groups of examinees, then the reading comprehension score reported probably has the same meaning for LD examinees with extended-time limits and NLD examinees with standard-time limits, at least in an internal sense (Rock, Bennett & Kaplan 1987; Rock, Bennett & Jirele, 1988 and Geisinger, 1994). The unit of analysis for the second part of the study was the passage-based testlet score obtained under each of the timing conditions. The following research questions were considered for the second part of this study:

5. Is the factor structure of the ITBS Reading Comprehension scores similar for the LD students under extended-time conditions and the NLD students under standard-time conditions?
6. Is the factor structure of the ITBS Reading Comprehension scores similar under extended-time conditions for both the LD and NLD students?
7. Is the factor structure of the ITBS Reading Comprehension scores similar for LD and NLD students under standard-time conditions?

Methods

For purposes of this study, a student without a learning disability (NLD) was defined as any student who was not identified by their school system as having a learning disability; they do not have an IEP (Individualized Education Program) that states there is a learning disability. A student with a learning disability (LD) was defined as any student whose primary disability is a learning disability as defined by the school system (self-contained or self-contained with integration or resource room learning disabled students) and shown in an IEP. LD students whose primary

disability is in a non-reading area (e.g., mathematics) are also included in this operational definition of LD for purposes of this study.

Students from two districts (A and B) took Form K, Level 12 of the Reading Comprehension test from the ITBS as part of their school's annual testing program. A total of 129 sixth-grade LD students made up the LD sample (LD4). Due to data collection inconsistencies, 61 of the LD students had both a standard-time score and extended-time score (LD3), and 68 had only an extended-time score. The majority of the LD students (83%) were participating in resource room programs.

The NLD comparison groups (n=409) were administered the ITBS Reading Comprehension test under both timing conditions at the same time as the LD groups (i.e., during normal ITBS administration dates). However, the results of the NLD groups from the two districts were not combined due to differences in directions used in each place during test administrations. Therefore, two separate NLD criterion groups were established. Criterion group 1 (NLD-A) from District A yielded a total of 235 out of 241 NLD students with usable data; the second criterion group, from District B (NLD-B), yielded a total of 162 out of 168 students with usable data.

Testing was conducted within two time periods during the same school year; District A conducted testing in February and District B tested in April. For both LD and NLD groups, testing was to resemble a standard administration with the exception of (a) the removal of statements in the directions regarding time limits and (b) the extension of time limits. If students under extended-time conditions needed more time to complete the test after the standard-time limit (40 minutes), they were asked to mark the last item answered on their answer folder and then were given additional 20-minute blocks of time until each student had a chance to complete the test.

Due to the practical constraints of scheduling within the school buildings, random assignment to treatments (i.e., timing condition) was not possible. Several methods were used to reduce the threats to the internal validity of the study: use of an instrument with sufficient floor and ceiling; use of the same modifications to the standardized testing directions for both groups; analysis of variance with LD and NLD students grouped based on Verbal Ability; and the use of difference scores

(extended-time score minus standard-time score) as the dependent measure to examine the changes in test performance.

Due to a lack of uniformity between the two districts in using the special directions to administer the ITBS Reading Comprehension test, a preliminary analysis of the similarities/dissimilarities in testing conditions was conducted. The resulting score distributions under each timing condition and the amount of time used for testing, in conjunction with interview data, were employed to determine if score data could be combined from the various locations within the LD and NLD groups. The primary findings of the combination analysis were that the NLD groups should be treated as separate groups and the majority of the LD students could be combined. (See Huesman (1999) for additional data that led to these conclusions.)

Results

The first part of the study examined the effects of extended time, compared to standard time, on test performance of LD and NLD students. Given the use of two NLD criterion groups, each hypothesis was tested twice, once with each NLD group. Unless stated otherwise, an $\alpha = .05$ was the level of probability considered for all statistical tests to be significant. Table 1 summarizes the test performance results in terms of the raw score means and the corresponding national grade equivalents for the two timing conditions by group. The average gain of approximately two raw score points for the NLD-A and LD3 groups is large: it represents an average growth on the national grade equivalent scale of six months for the LD students and four months for the NLD-A students. For the NLD-B students, the mean difference between standard-time and extended-time was less than one point. Extending time limits made little difference for this latter group of NLD students.

[Insert Table 1 about here]

Analysis of Difference Scores

Overall comparisons. A two-sample, independent, one-tailed t-test was used to test the hypothesis that the mean difference score for LD3 students is greater than the mean difference score for NLD students. Table 2 displays the mean difference scores (extended-time score minus standard-time score) of students with and without learning disabilities. The t-test was not significant for the NLD-A comparison but it was significant for the NLD-B comparison. The results support the hypothesis that students with learning disabilities make significantly larger gains on the ITBS Reading Comprehension Test under extended-time conditions than students without learning disabilities who received appropriate timing instructions. NLD students given instructions to take their time did not perform any differently than LD students under extended-time conditions.

[Insert Table 2 about here]

Unfortunately, the information needed to assess total elapsed time could only be collected for 45% (n=58) of the 129 LD students. Therefore, statements regarding comparisons of time across groups need to be tempered by this fact. For LD students with extended-time data, the average time spent on the ITBS Reading Comprehension test was 49 minutes. (The standard time limit was 40 minutes). The group of LD students for whose extended-time data was available (n=37), on average, used 16 additional minutes of testing time. Nearly two-thirds (65%; n=24/37) of these LD students finished within the first extra 20-minute block of time. With the exception of one LD student, who took a total of 85 minutes to finish the test, the remainder finished within the second 20-minute block of time. NLD-B students with extended-time data (67%; n=108/162) used, on average, 34 minutes to complete the test. The small group of NLD-B students (n=10) that utilized extended time used, on average, an additional seven minutes. For NLD-A students, 41.7% (n=98/235) used extra time, but only two students went into the second 20-minute block. No elapsed timing data was collected for this group.

Reading ability. In order to answer questions regarding the relationship between difference scores and reading ability, an analysis examining the shifts in the distributions of reading percentile

ranks was used to assess the impact of timing conditions for LD and NLD students of varying reading comprehension ability. Students were divided into six percentile rank groups based on their extended-time ITBS Reading Comprehension scores: ≤ 5, 6-24, 25-49, 50-74, 75-94, and 95-99, using midyear Iowa norms for the District A students and spring Iowa norms for the District B students. The extended-time score was used instead of the standard-time score because the former probably represents a more valid measure of reading comprehension (i.e., the effect of time would be less of a factor) for all students tested. The relationship between difference scores and reading comprehension ability was then examined via scatterplot analyses. Table 3 provides a summary of difference-score statistics by reading level for students with and without learning disabilities. Nearly 89% (n=54) of the students with learning disabilities were below the Iowa median on the ITBS Reading Comprehension test. As a group, 39.3% (n=24) of the LD3 students had difference scores greater than zero. Slightly more than one-third (n=55) of the NLD-B students were below the median on the ITBS Reading Comprehension test. As a group, however, only 10 NLD-B students had difference scores greater than zero. The majority of the NLD-A students (63%) were below the Iowa median on the ITBS Reading Comprehension test. As a group, 41.0% (n=96) of the NLD-A students had difference scores greater than zero. The majority of these 96 students (80.2%) were below the Iowa median on the ITBS Reading Comprehension test.

[Insert Table 3 about here]

A scatterplot analysis was completed to further examine the relationship between ITBS Reading Comprehension scores based on extended-time and difference scores by reading level (as defined earlier). Do poorer readers gain more than better readers under extended-time conditions, or is the difference much the same across reading levels? For students with learning disabilities, the average difference was fairly consistent across reading levels, though the small number of LD3 students above the median limits this generalization. The lowest-scoring LD3 students (percentile rank ≤5), in relation to their NLD peers, do make gains of note. This group represented a large number of the LD3 students (33%), and their gains under extended time represented a relatively large

difference in test performance as compared to either of the NLD groups at this reading level. Test performance for NLD-A students, as a group, increased across all reading levels except at the extremes of the distribution. For NLD-B students, the reverse was true: the vast majority did not benefit from extended-time, but a small number of above average students (n=7) made some gain under extended-time. See Huesman (1999) for scatterplot figures.

Gender. A general linear model with gender (male vs. female) and group (LD vs. NLD) as the between-subjects factors (i.e., two-way ANOVA) was utilized to examine the relationship between gender and difference scores. The results of this portion of the study are presented in Tables 4, 5, and 6. As a group, males and females showed similar changes in test performance: the two-way ANOVA showed no significant interaction effect for gender and group. Across all groups, females attained higher average ITBS Reading Comprehension scores than males under both timing conditions. The ratio of male to female students with learning disabilities used in the analysis of the interaction of gender with group was 2:1, compared to the nearly 1:1 ratio for the NLD students from both school districts. The average difference score did not vary by gender, and in fact, the reading comprehension levels of these system-identified female LD3 students was higher than their male LD3 counterparts, suggesting at least a less severe deficit in reading comprehension ability than would be predicted from the literature.

[Insert Tables 4, 5, and 6 about here]

Verbal ability. Is the relationship between the difference scores and verbal ability different within the LD3 and NLD groups. A general linear model with verbal ability and group (LD vs. NLD) as the between-subjects factors (i.e., two-way ANOVA) was utilized to answer this question. A student's verbal ability was categorized as below average, average, or above average based on his/her ITBS Vocabulary Iowa percentile rank, using midyear student norms for District A and spring student norms for District B. Below average verbal ability was defined as performance below the 25th percentile, average was defined as performance from at-or-above the 25th to at-or-below the 75th percentile, and above average was defined as above the 75th percentile. The ITBS Vocabulary test

was used as an indicator of verbal ability because no other single measure of verbal ability was available for each student. The results are presented in Tables 7, 8, and 9. Only one High-Verbal-Ability LD3 student was found. (Since a General Linear Models approach was conducted for the analysis of variance, the SPSS 8.0 algorithm provided an estimated standard error for this cell.) LD3 students in the low and average verbal ability groups appeared to have benefited equally under extended-time conditions. For NLD students, those in the average verbal ability group showed the largest gains, and for both NLD groups, those students in the high verbal ability group showed the smallest gains. None of these trends was statistically significant. The interaction between general verbal ability and group was not significant for either LD3 versus NLD-A or LD3 versus NLD-B.

[Insert Tables 7, 8, and 9 about here]

Stability of the Factor Structures

Due to the limited sample size of LD students, a descriptive approach was employed to address the second set of research questions. To assess the stability of the factor structures, composite variables (i.e. testlets) rather than individual items were used in the analysis of the factor structures because the relationship between individual dichotomously scored items is not linear (Rock et al., 1987 and Rock et al., 1988). The nonlinear relationship often results in identifying more factors than are really present: often items of similar difficulty group together whether they measure the same construct or not. The use of item parcels provides continuous scores that tend to have linear relationships with one another, and scores from parcels also provide more stable and reliable indicators of factors when comparing across populations. Seven passage-based testlets were formed for each timing condition, each based on the collection of items associated with each reading passage from the ITBS Reading Comprehension test. These composite variables were used as the unit of analysis for the score comparability section of the study.

A preliminary factor structure analysis was conducted as a first step to investigate the research questions. This approach utilized testlet summary statistics (average test performance and

reliability estimates) and examined the correlation coefficients between testlets under the two timing conditions. Estimates of testlet score reliability were calculated using coefficient alpha. Only estimates of the observed correlations between standard-time and extended-time scores were obtained; estimates of the disattenuated correlations could not be computed due to the correlated errors of standard-time and extended-time scores.

A Principal Components Analysis (PCA) based on the product-moment correlation matrices of testlet scores for each of the groups, by timing condition, was the next step. A comparison of the number of eigen values greater than one was used to examine the underlying dimensions of the common factor space of the ITBS Reading Comprehension test scores across groups and timing conditions. Comparisons involving the scores of LD students under standard-time conditions used the scores of LD3 students (n=61) and comparisons involving extended-time conditions used the scores of LD4 students (n=129). The extended-time summary results of LD3 students are also reported in order to assess the similarities/differences resulting from the addition of the LD students with only extended-time scores.

Descriptive Testlet Analysis. The results in Table 10 provide test performance information for each of the groups under each timing condition by the unit of analysis (i.e., passage-based testlets). Examination of the differences between testlet means in Table 10 reveals that increases in test performance under extended-time conditions occurred primarily toward the end of test (i.e., testlet 7) for the NLD-B group. For the NLD-A and LD groups, similar test performance increases started to occur as early as testlet 5.

[Insert Table 10 about here]

Table 11 contains the coefficient alpha estimates of testlet score reliability. The differences in test performance at the end of the exam are reflected in the estimated reliability coefficients as well. The observed reduced testlet score reliability estimates under extended-time conditions are primarily due to the decreased item-score variability and the accompanying decrease in overall testlet variability toward the end of the test. Given how standard-time and extended-time scores are

calculated, this should not come as a surprise: students who had not completed the test after the 40-minute mark would have had all responses after this mark coded as incorrect in the calculation of the standard-time score. Some of the responses after the 40-minute mark would be changed from incorrect to correct when calculating extended-time scores, therefore increasing the variability of the item scores and overall testlet variability.

[Insert Table 11 about here]

The analysis of the correlations between ITBS Reading Comprehension testlet scores for the two timing conditions was undertaken to determine whether the extension of time limits introduced new ability factors or reduced the influence of time in the standard administration. In either case, the correlations should be less than 1.0 (i.e., the relative rankings of the students should change). For NLD-B students, no correlations were very markedly different from 1.0. This would support the hypothesis that the two timing conditions measured essentially the same attributes for these students. However, for NLD-A students, the lowest correlation between standard-time and extended-time scores was 0.65 for testlet 7. For LD students, the smallest correlation also occurred at testlet 7, (0.47).

Principal Components Analysis. Tables 12 and 13 present the eigenvalues and the differences between successive components in order to assess the magnitude of the difference between the first and second eigenvalues. The factor structure analyses of the ITBS Reading Comprehension scores under the two timing conditions mirrors the test performance results: both the NLD-A and the LD group showed slight evidence of a second factor under standard-time conditions (see Table 12), though the presence of the second factor was strongest for the LD group. This second factor was not apparent for the NLD-B group, which worked at a more normal rate (see Table 13). The observed testlet correlations between standard-time and extended-time scores also verified these observations. The correlations for the last testlet were quite different from 1.0 for the NLD-A and LD3 groups compared to the NLD-B group. This pattern was also corroborated by the observed testlet reliability estimates (see Table 11). Under extended-time conditions, the indication of the

second factor diminished for the NLD-A group, though evidence remained of a second factor for the LD group.

[Insert Tables 12 & 13 about here]

Implications

An important finding of this research is that the amount of extra time needed by LD students varies a great deal, ranging from no extra time for some students to an additional 45 minutes for others. Thus, the use of an arbitrary, universal rule for IEPs that would permit twice the time limit, or some other "standard" value for all LD students, would not be appropriate for IEP writings. Runyan (1991a) pointed out, "...students with learning disabilities have varied rates of processing printed information, and therefore a fixed amount of extra time for all students with learning disabilities on standardized tests may not be appropriate" (p. 107). Perhaps in the future, when student's IEPs are being developed, a measure of their rate of processing information can be incorporated to estimate the amount of time needed on standardized achievement tests, depending on the level and type of disability. Such an approach would take into account the variability found within the LD population and prevent some students from getting either too much or too little time.

The extension of time limits is only one of the many types of accommodations that could be given to students with learning disabilities. Some students may need only portions of a test read verbatim (if this is appropriate), while others may need a reader for the entire test, and still others may only need extra time without a reader. It is not the case that all students with learning disabilities have low reading levels: for the LD3 group, 11% of the 61 (see Table 3) students were above the Iowa median on a test they took at grade level under extended-time conditions. This suggests that all LD students do not need special accommodations (e.g., reading of test materials) when taking standardized achievement tests. It seems unreasonable to make blanket rules regarding testing accommodations, given the variation in reading ability that was found in this study.

The effect of testing directions on student performance became apparent when comparing the performance levels of two groups of NLD students. NLD-A students told to work at a slow and careful pace, where time was not a factor, made significant gains under extended-time conditions, beyond the levels of the grade 6, 1997 cohort from their school district (national grade equivalents, 6.4 vs. 5.8, respectively). NLD-B students, told to work at a normal rate, did not make significant gains under extended-time conditions (i.e., work as quickly as possible, but not so fast as to not do your best work). They most likely interpreted this to mean, "Work as though you are being timed." NLD-B test performance, under extended-time conditions was very similar to the levels of the grade 6, 1997 cohort from their school district (national grade equivalents, 8.4 vs. 8.7, respectively).

There are implications for removing statements from test directions regarding time and for extending time limits. The notion of extending time limits so that all students have a chance to finish the test would result in the need to restandardize ITBS Reading Comprehension test scores, especially, if students are told to work slowly. The evidence was clear that the scores obtained under standard-time conditions differed in meaning for NLD-A and NLD-B students. The evidence was less clear that the scores of the NLD and LD students under extended-time conditions had similar meaning, in terms of the underlying factor structure of the scores. Perhaps, the evidence of the second factor under both timing conditions for the LD students is due to all students taking an on-level test (i.e., a difficulty factor). If the goal is to include all students in the assessment process and be able to compare scores obtained under similar testing conditions, it may make more sense to give all students extended time using methods like those employed in this study (i.e., in 20-minute blocks of time). This approach might make the practical impact of extended time on the school day schedule for teachers and administrators more manageable, while individualizing the amount of time needed for each student (LD or NLD).

The observed increase in test performance for the LD group and the corresponding lack of a significant increase for a comparison NLD group (NLD-B students) under extended-time conditions has been found by other researchers (Alster, 1997; Harker, 1991; Hill, 1984; Runyan, 1991a, 1991b).

Exceptions were Halla (1988) and Munger and Lloyd (1991). The results from the two NLD comparison groups in this study, based on different testing directions, provide support for both of these previous findings. The NLD-B group working at a normal rate, corroborated the first set of results, but the NLD-A group, which was instructed to work carefully and take its time, tended to make gains as great as their LD counterparts. The exception occurred for the lowest scoring LD students, who tended to make larger gains in relation to their NLD (District A or B) peers at the same reading level.

If the amount of testing time was not an issue for students with learning disabilities, it would be reasonable to expect results for them like those of the NLD-B group: difference scores near zero; similar testlet means, and reliability estimates; and stable factor structures across timing conditions. For the NLD-B students, there was little difference in the underlying factor structure: one prominent principal component accounted for the majority of the variance, regardless of timing conditions. When NLD-A students were directed to use as much time as needed, the opposite resulted: difference scores greater than zero; dissimilar testlet means, and reliability estimates toward the end of the test; and factor structure differences. A two-factor model appeared to fit their data better under standard-time conditions, whereas, under extended-time conditions, a one-factor model appeared to fit better. This second factor was hypothesized to be a rate of work or speededness factor. Again, given their instructions, this is evidence that the extension of time limits introduced an additional factor for the NLD-A students that was not present in the NLD-B group. The NLD-A results were similar to those of the LD students, who were probably working at a more "normal" rate. Test performance of LD students significantly increased under extended-time conditions, and the factor structure changed, though not to the same extent as that of the NLD-A group. The additional factor under standard-timing conditions was hypothesized to be a speededness factor for the LD students.

For LD students, the second factor diminished under extended-time conditions, which may be evidence that the extension of time limits reduced an important attribute present in the test administration for LD students (i.e., speededness). This finding could be explained by the processing

speed deficits that have been found with students with learning disabilities. The extra testing time probably compensates for the slower speed of processing and removes an irrelevant source of difficulty from most scores of standardized tests involving reading. Perhaps if subject selection procedures could have been more rigorous and only students with reading disabilities had been selected for this study, the observed reduction in the second eigen value might have been even greater. Another related possibility is that the second factor represented, in part, a difficulty factor related to test level. Since all LD students took an on-level test, the extension of time limits may have only diminished, but not eliminated, the second factor from the underlying factor structure of the ITBS Reading Comprehension test scores (see Table 12).

It is difficult to assess the impact of the severity of the student's disabilities since access to IEPs was restricted. But the high proportion of the LD students in resource room type programs suggests that most had less severe disabilities. The impact of multiple disabilities (e.g., behavioral disorders, etc.) on the results cannot be assessed. The average score under extended-time conditions increased for LD students, but it was still less than the average standard-time score for either the NLD-A or NLD-B groups. Therefore, it appears that the extension of time limits removed or reduced an irrelevant source of difficulty for students with learning disabilities. The large proportion of low to moderate achieving students in the NLD-A group may have also been aided by the directions to work carefully and take their time because, as a group, their average difference score was nearly the same as that of the LD students.

For the LD students, the lack of information on the type of disability, the severity of the disability, and the presence of multiple disabilities confound the interpretation of the observed results. This lack of information also limits the generalizability of the results to elementary-grade students with a primary label of learning disability rather than to the narrower group of those with reading disabilities. In addition, the use of school-identified samples of students with learning disabilities may be biased with regard to gender. If females are underrepresented or if those identified have more or less severe deficits than most LD females, the results may not be generalizable to the general

elementary-grade LD population. However, to the extent that similar guidelines were followed for assessing and identifying LD students in Districts A and B, the results should be generalizable to a broader LD population.

Few studies have attempted to study the score comparability of LD and NLD students under extended-time conditions for elementary achievement tests. The reasons for this no doubt lie in the difficulty of obtaining appropriate sample sizes of LD students and obtaining access to IEPs and psychological reports in order to determine types of disabilities and the level of their severity. Future studies should involve more researcher control of the entire testing process: distributing special directions, conducting workshops for test administrators on appropriate special test administration procedures, and using multiple trained observers in the classrooms during testing. These procedures would help ensure uniform test administration and better data collection across buildings, which should increase the power of the statistical tests and make interpretation of the results clearer and more generalizable.

The effects of different pacing directions on test performance should be examined across districts and buildings, in order to study the generalizability of this finding across groups of varying achievement. For example, did lower achieving NLD students (NLD-A) benefit more, on average, when told to take their time and work carefully and would higher achieving NLD students (NLD-B) also benefit, on the average, from this type of direction and obtain even higher scores? Runyan's (1991a, 1991b) directions to students were similar to those given to the NLD-B students, that is, no mention of timing was made and students were told to work at a normal pace. Based on the results found in this study, it may be the case that NLD students who work at a normal rate, thinking they are being timed, may not have reached their maximum scores, while those NLD students working under a different rate of work, where time is not a factor, tended to show greater improvement and more valid reading scores.

The selection of LD students with only reading difficulties would have also aided in the interpretation of the observed results. Is it the case, that learning disabled students of all types benefit

equally from extended time on reading tests, or does it vary by the type of learning disability (e.g., reading, math, listening, etc.)? In future studies, researchers need access to student records to obtain such data so that large and relevant samples of LD subgroups can be included.

Although difficult, the need for validity studies using LD students is great. Due to legal mandates and social trends, information on the comparability of scores resulting from non-standard assessments must be obtained. Evidence of the validity of achievement test scores arising from accommodated conditions for disabled groups, in relation to their non-disabled peers, must become part of the validity evidence that is now gathered regularly by testing programs and test publishers.

Table 1. ITBS Reading Comprehension Mean Raw Scores and National Grade Equivalents for Standard-Time and Extended-Time Conditions

	Mean Raw Score		National Grade Equivalent		National Grade Equivalent
Group	Standard	Extended	Standard	Extended	Difference
LD3	17.5	19.7	4.60	5.21	0.61
NLD-A	24.2	26.2	6.24	6.62	0.38
NLD-B	32.0	32.3	8.30	8.39	0.09

Notes: LD3 = LD Students with both standard-time and extended-time scores (n=61)
NLD-A = NLD students from District A (n=235)
NLD-B = NLD students from District B (n=162)
National Grade Equivalent values estimated via linear interpolation

Table 2. Comparison of ITBS Reading Comprehension Mean Difference Scores for LD and NLD Students

Group	N	Mean Difference	Standard Deviation	t-value	df	p
LD3	61	2.16	3.11			
NLD-A	235	1.94	2.85	-0.54	294*	0.592
NLD-B	162	0.26	1.22	-4.65	67**	0.000

Notes: NLD-A = NLD students from District A (n=235)
NLD-B = NLD students from District B (n=162)
LD3 = LD Students with both standard-time and extended-time scores (n=61)
*equal variances assumed **equal variances not assumed

BEST COPY AVAILABLE

Table 3. Mean Difference Scores for LD and NLD Students by Reading Ability Groups

Reading Iowa Percentile Rank	Group	N	Mean Difference	Standard Deviation
≤5	NLD-A	24	0.33	0.92
	NLD-B	4	0.00	0.00
	LD3	20	1.65	2.87
>5 - <25	NLD-A	63	2.25	2.65
	NLD-B	23	0.09	0.42
	LD3	20	2.20	3.27
≥25 - <50	NLD-A	61	2.33	2.88
	NLD-B	28	0.07	0.38
	LD3	14	2.71	3.43
≥50 - <75	NLD-A	55	2.11	3.23
	NLD-B	58	0.64	1.96
	LD3	5	2.00	2.74
≥75 - <95	NLD-A	26	1.62	3.24
	NLD-B	38	0.03	0.16
	LD3	2	3.50	4.95
≥95	NLD-A	6	1.00	2.45
	NLD-B	11	0.00	0.00
	LD3	.	.	.
Total	NLD-A	235	1.94	2.85
	NLD-B	162	0.26	1.22
	LD3	61	2.16	3.11

Table 4. Mean Difference Scores and Standard Deviations by Group and Gender

Group	Gender	N	Mean Difference	Standard Deviation
NLD-A	Female	117	1.90	2.83
	Male	118	1.98	2.87
NLD-B	Female	82	0.18	1.01
	Male	80	0.34	1.41
LD3	Female	20	2.30	3.11
	Male	41	2.10	3.14

Notes: NLD-A = NLD students from District A (n=235)
NLD-B = NLD students from District B (n=162)
LD3 = LD students with both standard and extended-time scores (n=61)

Table 5. ANOVA Summary Table for Group (LD3 vs. NLD-A) and Gender

Source of Variation	df	Mean Square	F Ratio	p
Group	1	2.93	0.35	0.56
Gender	1	0.15	0.02	0.89
Group X Gender	1	0.91	0.11	0.74
Error	292	8.74		
Total	295			

Table 6. ANOVA Summary Table for Group (LD3 vs. NLD-B) and Gender

Source of Variation	df	Mean Square	F Ratio	p
Group	1	151.71	40.52	0.00
Gender	1	0.02	0.01	0.94
Group X Gender	1	1.29	0.34	0.59
Error	219	3.74		
Total	222			

Table 7. Mean Difference Scores and Standard Deviations by Group and Verbal Ability

Group	Verbal Ability	N	Mean	Standard Deviation
NLD-A	High	35	1.06	2.62
	Average	98	2.36	3.19
	Low	100	1.79	2.41
	Total	233	1.92	2.82
NLD-B	High	64	0.02	0.13
	Average	74	0.53	1.76
	Low	24	0.08	0.41
	Total	162	0.26	1.22
LD3	High	1	5.00	.
	Average	16	2.25	3.02
	Low	44	2.07	3.18
	Total	61	2.16	3.11

Notes: NLD-A = NLD students from District A (n=235)
NLD-B = NLD students from District B (n=162)
LD3 = LD Students with both standard-time and extended-time scores (n=61)

Table 8. ANOVA Summary Table for Group (LD3 vs. NLD-A) and Verbal Ability

Source of Variation	df	Mean Square	F Ratio	p
Group	1	14.92	1.82	0.18
Verbal Ability	2	4.50	0.55	0.58
Group x Verbal Ability	2	7.52	0.92	0.40
Error	288	8.22		
Total	293			

Table 9. ANOVA Summary Table for Group (LD3 vs. NLD-B) and Verbal Ability

Source of Variation	df	Mean Square	F Ratio	p
Group	1	65.36	14.18	0.00
Verbal Ability	2	4.56	1.23	0.29
Group x Verbal Ability	2	4.88	1.32	0.27
Error	217	3.70		
Total	222			

Table 10. Testlet Raw Score Statistics for ITBS Reading Comprehension Test by Group and Timing Condition

		NLD-A		NLD-B		LD3		LD4
		Standard	Extended	Standard	Extended	Standard	Extended	Extended
Testlet	k*	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
1	8	5.55 (1.82)	5.55 (1.82)	6.48 (1.55)	6.48 (1.55)	4.59 (2.10)	4.59 (2.10)	4.38 (2.14)
2	4	2.16 (1.11)	2.16 (1.11)	2.62 (1.14)	2.62 (1.14)	1.74 (1.06)	1.74 (1.06)	1.68 (1.07)
3	5	3.34 (1.39)	3.34 (1.39)	3.91 (1.31)	3.91 (1.31)	2.43 (1.55)	2.51 (1.56)	2.25 (1.49)
4	6	3.60 (1.50)	3.62 (1.48)	4.42 (1.22)	4.42 (1.22)	2.82 (1.53)	2.90 (1.42)	2.81 (1.48)
5	7	4.29 (2.19)	4.54 (2.02)	5.43 (1.73)	5.43 (1.73)	2.72 (2.26)	3.11 (1.98)	3.16 (1.84)
6	7	3.20 (2.31)	3.83 (2.04)	5.09 (1.67)	5.15 (1.59)	1.93 (2.06)	2.57 (1.92)	2.67 (1.68)
7	7	2.08 (2.17)	3.12 (1.96)	4.04 (1.87)	4.23 (1.68)	1.30 (1.55)	2.26 (1.71)	2.36 (1.62)
Overall	44	24.21 (8.90)	26.15 (8.57)	31.99 (7.84)	32.25 (7.75)	17.52 (8.63)	19.69 (8.58)	19.00 (7.17)

* Notes: k = number of items per testlet
NLD-A = NLD students from District A (n=235)
NLD-B = NLD students from District B (n=162)
LD3 = LD students with both standard-time and extended-time scores (n=61)
LD4 = LD students with extended-time scores (includes LD3), (n=129)

BEST COPY AVAILABLE

Table 11. Coefficient Alpha Reliability Estimates for the ITBS Reading Comprehension Testlet Scores by Group and Timing Condition

	NLD-A		NLD-B		LD3		LD4
Testlet	Standard	Extended	Standard	Extended	Standard	Extended	Extended
1	0.60	0.60	0.60	0.60	0.66	0.66	0.68
2	0.33	0.33	0.46	0.46	0.26	0.26	0.25
3	0.59	0.59	0.65	0.65	0.60	0.60	0.55
4	0.56	0.54	0.52	0.52	0.50	0.39	0.43
5	0.78	0.73	0.71	0.71	0.80	0.67	0.60
6	0.81	0.72	0.62	0.58	0.79	0.66	0.51
7	0.81	0.66	0.63	0.52	0.65	0.56	0.48
Overall	0.90	0.89	0.89	0.89	0.89	0.88	0.85

Notes: NLD-A = NLD students from District A (n=235)
NLD-B = NLD students from District B (n=162)
LD3 = LD students with both standard-time and extended-time scores (n=61)
LD4 = LD students with extended-time scores (includes LD3), (n=129)

Table 12. Eigenvalues of Product-Moment Correlation Matrices Based on Testlet Scores: NLD-A vs. LD

	NLD-A: Standard			NLD-A: Extended			LD3: Standard			LD4: Extended		
Rank of Eigenvalue	Eigenvalue	%	D	Eigenvalue	%	D	Eigenvalue	%	D	Eigenvalue	%	D
1	3.48	49.7	2.44	3.65	52.1	2.74	3.52	50.2	2.27	3.30	47.2	2.24
2	1.04	14.8	0.36	0.91	13.0	0.24	1.25	17.9	0.38	1.06	15.1	0.30
3	0.68	9.7	0.09	0.67	9.6	0.10	0.87	12.5	0.39	0.76	10.9	0.16
4	0.59		0.11	0.57		0.09	0.48	6.8	0.10	0.60	8.6	0.08
5	0.48		0.05	0.48		0.11	0.38		0.12	0.52		0.03
6	0.43		0.11	0.37		0.01	0.26		0.02	0.49		0.21
7	0.32			0.36			0.24			0.28		

Notes: NLD-A = NLD students from District A (n=235)
LD3 = LD students with both standard-time and extended-time scores (n=61)
LD4 = LD students with extended-time scores (includes LD3), (n=129)
% = Percentage of observed variance accounted for
D = Difference between successive eigen values

BEST COPY AVAILABLE

Table 13. Eigenvalues of Product-Moment Correlation Matrices Based on Testlet Scores: NLD-B vs. LD

	NLD-B: Standard			NLD-B: Extended			LD3: Standard			LD4: Extended		
Rank of Eigenvalue	Eigenvalue	%	D	Eigenvalue	%	D	Eigenvalue	%	D	Eigenvalue	%	D
1	3.90	55.7	3.06	4.00	57.1	3.20	3.52	50.2	2.27	3.30	47.2	2.24
2	0.84	12.0	0.15	0.80	11.4	0.13	1.25	17.9	0.38	1.06	15.1	0.30
3	0.69	9.8	0.14	0.67	9.6	0.12	0.87	12.5	0.39	0.76	10.9	0.16
4	0.55		0.11	0.55		0.16	0.48	6.8	0.10	0.60	8.6	0.08
5	0.44		0.13	0.39		0.08	0.38		0.12	0.52		0.03
6	0.31		0.03	0.31		0.02	0.26		0.02	0.49		0.21
7	0.28			0.29			0.24			0.28		

Notes: NLD-B = NLD students from District B (n=162)
LD3 = LD students with both standard-time and extended-time scores (n=61)
LD4 = LD students with extended-time scores (includes LD3), (n=129)
% = Percentage of observed variance accounted for
D = Difference between successive eigen values

BEST COPY AVAILABLE

References

Ackerman, P. T., Dykman, R. A., & Peters, J. E. (1977). Learning-disabled boys as adolescents: Cognitive factors and achievement. Journal of the American Academy of Child Psychiatry, 16(2), 297-313.

Alster, E. H. (1997). The effects of extended time on algebra test scores for college students with and without learning disabilities. Journal of Learning Disabilities, 30(2), 222-227.

Anderson, K. G. (1997). Gender bias and special education referrals. Annals of Dyslexia, 47, 151-162.

Badian, N. A. (1996). Dyslexia: A validation of the concept at two age levels. Journal of Learning Disabilities, 29(1), 102-112.

DeFries, J. C. (1989). Gender ratios in children with reading disability and their affected relatives: A commentary. Journal of Learning Disabilities, 22(9), 544-545.

Dodd, J. M., Griswold, P. E., Smith, G. H. & Burd, L. (1985). A comparison of learning disabled and other children on the ability to make functional time estimates. Child Study Journal, 15(3), 189-197.

Geary, D. C., & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. Developmental Psychology, 27(3), 398-406.

Geisinger, K. F. (1994). Psychometric issues in testing students with disabilities. Applied Measurement in Education, 7(2), 121-140.

Halla, J. W. (1988). A psychological study of psychometric differences in graduate record examinations general test scores between learning disabled and non-learning disabled adults (Doctoral dissertation, Texas Tech University, Lubbock). Dissertation Abstracts International, 49, 3341A.

Harker, J. K. (1991). A comparison of achievement test performance under silent reading and reading plus listening modes of administration (Doctoral dissertation, The University of Iowa, Iowa City). Dissertation Abstracts International, 52, 4304A.

Hasselbring, T. S., Goin, L. I., & Bransford, J. D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. Focus On Exceptional Children, 20(6), 2-7.

Hayes, F. B., Hynd, G. W., & Wisenbaker, J. (1986). Learning disabled and normal college students' performance on reaction time and speeded classification tasks. Journal of Educational Psychology, 78(1), 39-43.

Hill, G. A. (1984). Learning disabled college students: The assessment of academic aptitude (Doctoral dissertation, Texas Tech University, Lubbock). Dissertation Abstracts International, 46, 642B.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1994). Iowa Tests of Basic Skills Forms K and L. Chicago: The Riverside Publishing Company.

Huesman, R. L. (1999). The validity of ITBS reading comprehension test scores for learning disabled and non learning disabled students under extended-time conditions (Doctoral dissertation, The University of Iowa, Iowa City). Dissertation Abstracts International, 60, 06A.

Individuals with Disabilities Education Act of 1991, 20 U.S.C 1400 et seq (1991).

Individuals with Disabilities Education Act Amendments of 1997, P.L. 105-17 (1997).

Kulak, A. G. (1993). Parallels between math and reading disability: Common issues and approaches. Journal of Learning Disabilities, 26(10), 666-673.

McDonnell, L. M., McLaughlin, M. J. & Morison, P. (1997). The diversity of students with disabilities. In L. M. McDonnell, M. J. McLaughlin, & P. Morison (Eds.), Educating one and all (pp. 68-112). Washington, DC: National Academy Press.

Munger, G. F. & Loyd, B. H. (1991). Effect of speededness on test performance of handicapped and nonhandicapped examinees. Journal of Educational Research, 85(1), 53-57.

Nass, R. D. (1993). Sex differences in learning abilities and disabilities. Annals of Dyslexia, 43, 61-77.

National Center on Educational Outcomes. (August, 1996). A review of the literature on testing accommodations for students with disabilities (Minnesota Assessment Report No. 9). Minneapolis, MN: University of Minnesota.

Perlman, C., Borger, J., Collins, C., Elenbogen, J., & Wood, J. (1996, April). The effect of extended time limits on learning disabled students' scores on standardized reading tests. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. Applied Measurement in Education, 7(2), 93-120.

Phillips, S. E. (1996). Legal defensibility of standards: Issues and policy perspectives. Educational Measurement: Issues and Practice, 15(2), 5-13, 19.

Rock, D. A., Bennett, R. E., & Jirele, T. (1988). Factor structure of the Graduate Record Examinations General Test in handicapped and nonhandicapped groups. Journal of Applied Psychology, 73(3), 383-392.

Rock, D. A., Bennett, R. E., & Kaplan, B. A. (1987). Internal construct validity of a college admissions test across handicapped and nonhandicapped groups. Educational and Psychological Measurement, 47, 193-205.

Runyan, M. K. (1991a). The effect of extra time on reading comprehension scores for university students with and without learning disabilities. Journal of Learning Disabilities, 24(2), 104-108.

Runyan, M. K. (1991b). Reading comprehension performance of learning-disabled and non-learning disabled college and university students under timed and untimed conditions (Doctoral dissertation, University of California, Berkeley). Dissertation Abstracts International, 52, 2875A.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut longitudinal study. Journal American Medical Association, 264(8), 998-1002.

SPSS for Windows Release 8.0 [Computer Software]. (1997). Chicago, Illinois: SPSS Inc.

Vogel, S. A. (1990). Gender differences in intelligence, language, visual-motor abilities, and academic achievement in students with learning disabilities: A review of the literature. Journal of Learning Disabilities, 23(1), 44-52.

Willingham, W., Ragosta, M., Bennett, R., Braun, H., Rock, D., & Powers, D. (1988). Testing handicapped people. Boston: Allyn and Bacon.